

# Задача машинного обучения.

Сформулируем задачу ML следующим образом: имеются 2 множества:

- множество  $X$  объектов,
- множество  $Y$  ответов.

Предполагается, что существует функциональная зависимость  $f : X \rightarrow Y$  между объектами и ответами, но она неизвестна. Известна лишь совокупность  $S$  пар вида (объект, ответ), называемая обучающей выборкой (training sample):

$$S = \{(x_i, y_{x_i} = f(x_i)) \in X \times Y | i = 1, \dots, D\}$$

Задача ML - найти приближенный функции  $f$  путем построения аппроксимирующей функции  $a_s : X \rightarrow Y$ , такой что  $\forall x_i \in X \quad \exists a_s(x_i) \approx f(x_i)$

## Способ описания объектов. Признаковое описание

Зададим множество  $F$  признаков объектов (features). Каждому признаку сопоставляется множество  $D_i$  значений этого признака:

$$f_j : X \rightarrow D_j, j = 1, \dots, n$$

Каждый признак  $f_i$  имеет определенный тип. Приведем примеры некоторых из таких типов и соответствующие им множества значений  $D_i$

- бинарный:  $D_i = \{-1, 1\}$  или  $\{0, 1\}$ ,
- номинальный:  $D_i$  – конечное множество,
- порядковый:  $D_i$  – конечное линейно упорядоченное множество,
- количественный:  $D_i = \mathbb{R}$  (множество действительных чисел).

Вектор  $(f_1(x), \dots, f_n(x))$  представляет собой признаковое описание объекта  $x$ . В общем случае для множества объектов  $X$  мы получим матрицу признаков объектов:

$$F = \|f_i(x_j)\|_{n \times D} = \begin{pmatrix} f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_D) & \cdots & f_n(x_D) \end{pmatrix}$$

## Как задаются ответы. Типы задач.

Способ того каким образом задается множество ответов позволяет нам провести классификацию задач.

**Задачи обучения с учителем (supervised learning):** заданы "ответы учителя"  $y_i = y(x_i)$  на обучающих  $x_i$ , т.е. существует обучающая выборка:

- задача классификации( classification,  $Y$  class labels):

- $Y = \{-1, +1\}$  2 класса (binary classification)
- $Y = \{1, \dots, M\}$  много классов (multiclass .)
- $Y = \{0, 1\}^M$  пересекающиеся классы (multilabel .)
- задача регрессии (regression):  $Y = \mathbb{R}$  или  $Y = \mathbb{R}^m$
- задача ранжирования (ranking, learning to rank):  $Y$  конечное упорядоченное множество

**Задачи обучения без учителя (unsupervised learning):** (unsupervised learning): ответов нет, необходимо что-то делать с самими объектами

## Минимизация эмпирического риска

Алгоритм обучения (learning algorithm) представляет собой алгоритм нахождения по обучающей выборке  $S$  такой аппроксимирующей функции  $a_s$ , которая обладает свойствами оптимальности:  $a_s$  должно как можно лучше приближать исходную неизвестную функцию  $f : X \rightarrow Y$  на всем  $X$ .

Для точного описания свойств оптимальности алгоритмов обучения используется понятие функции потерь (loss function), которая сопоставляет паре  $(a_s, x)$ , где  $x \in X$  функцию  $L(a_s, x)$ , выражающее величину ошибки аппроксимации  $a_s$  на объекте  $x \in X$ .

Определим понятие эмпирического риска, как функционал качества, характеризующий среднюю ошибку алгоритма на  $a_s$  выборке  $S$ .

$$Q(a, S) = \frac{1}{D} \sum_{i=1}^D L(a_s, x)$$

Метод минимизация эмпирического риска заключается в том, чтобы найти алгоритм, доставляющий минимальное значение функционалу эмпирического риска.

$$a = \arg \min Q(a, S)$$

Основное достоинство рассмотренного метода заключается в том, что это конструктивный и универсальный подход, позволяющий сводить задачу обучения к задачам численной оптимизации.

## Предсказательная модель

Как правило, для решения задачи построения функции  $a_s : X \rightarrow Y$  по обучающей выборке  $S$  выбирается некоторая модель обучения, состоящая из двух компонентов:

1. Первой компонентой модели обучения является функция, которая выбирается из параметрического семейства функций

$$a : X \times W \rightarrow Y$$

где  $W$  – множество, элементы которого называются параметрами. Искомая

функция  $a_S$  ищется в виде

$$a_S(x) = a(x, w),$$

где  $w$  – фиксированный параметр. Функцию  $a_S$  иногда называют **предсказательной моделью (predictive model)**.

2. Другой компонентой модели обучения является алгоритм обучения, который представляет собой алгоритм поиска такого значения  $w$ , для которого функция  $a_S$ , обладает некоторыми свойствами оптимальности.

### Пример:

Рассмотрим линейную предсказательную модель В линейной модели множество  $W$  параметров имеет вид  $\mathbb{R}^n$ , где  $n$  – число признаков объектов, т.е. каждый параметр  $w$  представляет собой вектор действительных чисел  $w = (w_1, \dots, w_n)$ , и

- в задачах регрессии и ранжирования  $Y = \mathbb{R}$ , и

$$a(x, w) = \langle x, w \rangle = \sum_{i=1}^n x^i w_i$$

- в задачах классификации  $Y = \{-1, 1\}$ , и

$$a(x, w) = \text{sign}(\langle x, w \rangle)$$

где  $\text{sign}$  – функция знака, она сопоставляет неотрицательным числам значение 1, а отрицательным – значение  $-1$ .